

Rating Credibility of Sources for Profiling Risk and Business Context of Service Requests

Rebecca Copeland, Noel Crespi
Institute Mines-Telecom, Telecom SudParis, Paris, France

Abstract— Attributes credibility, trustiness and accuracy are often considered, but not incorporated in decision making algorithms. In this paper a credibility-based risk/business context model is proposed, to profile communication service requests using sources' credibility, as well as observed intensity and customizable policy-based prioritization. The paper proves that this Credibility-based approach can cope with the complexity and uncertainty of context models, which Bayes probabilities cannot, and with degrees of discordance and uncertainty, which DST cannot. The paper provides reviews of concepts and techniques to determine credibility from its components and aggregate corroborated credibility within observations and ultimately key factors of the model. The paper tests and proves the effectiveness of the modelled credibility, intensity and policy, and shows that incorporating Credibility is a better way to make informed decisions.

Keywords—Context; Credibility; Dempster rule; DST; Belief function; Bayes; AHP; Entropy; Discordance; OWA; SAW; WPM;

I. INTRODUCTION

Establishing users' context at the point of requesting a service is highly desirable for increasing number of functions, such as establishing business priority, selective funding, optimizing access choice and more. However, while cyber-crime is rapidly rising, the greatest benefit would come from early detection of risks at the time of requesting a service, before granting access to sensitive data and scarce resources. This type of context analysis derives information not only from the request's details, but also from users' situation and behavior. The context model hinges on situational aspects such as 'Integrity' and 'Urgency'. Attributes are determined by indirect information and inference procedures that involve a high level of uncertainty. Therefore, as much information as can be gathered should be utilized to assemble context details and enhance the decision reliability.

Behavioral context must carry a high level of confidence that the 'verdict' is accurate enough to act upon it. The plausibility of evidence is assessed by either 'metric' means or 'probabilistic' predictions. Credibility of metric observations is based on the specifications of sensitivity or failure rates of the measuring instruments and sensors. However, context information is not directly 'sensed', but is inferred from 'hints' and indirect indications, so gauging credibility of the whole process of extracting information is essential.

Probabilistic models rely on previous occurrences of context attributes in statistical data. Many Bayes algorithms base their forecast entirely on generated 'training data sets', which are never updated by real results. While creating such training data for a small number of variables is feasible, this is not the case for the much larger numbers of attributes that are

needed for evaluating situational context. Additionally, Bayes models stipulate attributes independence, which is impossible to achieve for context clues and supplemental evidence. Hence a new approach is sought, to cope with uncertainty and contradiction, while producing definitive decisions, according to preferred policies. This requires obtaining fresh observations with their intensity and degrees of support or disagreement, and robust mechanisms of inference and aggregation.

This paper proposes an approach that incorporates sources' *Credibility* in the context evaluation. This approach has been implemented in an enterprise Business & Risk (eBCR) Context model, which was first described in [20],[21]. In this paper, credibility structure and rating are proposed, with methods of aggregating context scores using credibility, intensity and prioritization. In Section II, types of aggregation are reviewed. In Section III, the credibility approach is explained. Section IV computing Credibility procedures are discussed. In Section V, alternatives approaches (Entropy, Bayes and DST) are investigated. In Section VI, the effectiveness of model components is analyzed. In VII related work, and in VIII, conclusions are given.

II. THE CREDIBILITY-BASED APPROACH

A. Definition of Credibility and its Components

Credibility components include Confidence, Accuracy and Precision (CAP), which are terms that are often confused. An observation that is precise is not necessarily accurate- stating that there are 58 red sweets in the jar is *precise*, but perhaps not *accurate*, and without means of counting the sweets, the *confidence* in the estimate is low. Credibility concepts, as shown in **Figure 1**, are defined in the International Vocabulary of Metrology [4], stating that "measurement '*accuracy*' should not be used for measurement '*trueness*'" and "measurement '*precision*' should not be used for measurement '*accuracy*'".

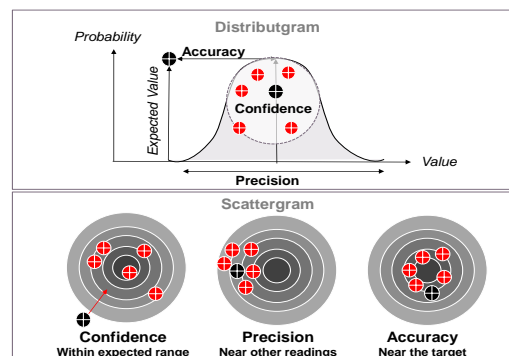


Figure 1. Credibility's Confidence, Precision and Accuracy definitions

Confidence (evidential reliability) is defined in [4] as the “coverage probability”, i.e. the likelihood that a data item is contained within a specified interval. Confidence is “close agreement between *measured* quantity value and a *true* quantity value”, which is theoretically replicated by infinite number of tests. Confidence that a data item is within the ‘coverage-range’ is interpreted as the combined effect of reliability, availability, manageability, efficiency of data retrieval and the level of data changeability of the sources.

Accuracy (correctness and data integrity) is defined in [4] as “closeness to a true or an expected value”. Accuracy is often considered as correctness (positive/negative true/false status), but such determination is not available for service requests. Therefore, accuracy is derived from estimated properties of error management and error prevention procedures, e.g. secure data input or extensive data audit and validation procedures.

Precision (level of details) is defined in [4] as “closeness of agreement between indications obtained by replicate measurements”. Precision includes Resolution, Proximity and Inclusion Zone. Distance proximity is crucial to spatial attributes, but temporal proximity affects many assertions, where their relevance is determined by elapsed time.

B. Credibility Component Structure

Each source is assessed by credibility components, in three layers: Measures, Traits and Elements, as shown in Figure 2. This allows for source characteristics, such as reliable retrieval, tamper-proof storage or data freshness, to be assessed individually. Procedures of data entry, storing and retrieval are also considered as atomic elements.

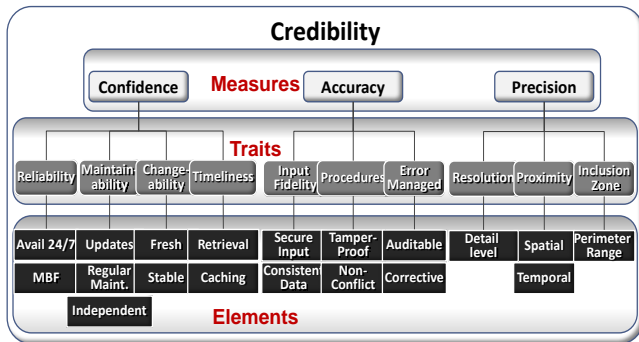


Figure 2. Credibility components structure

Measures include Confidence, Accuracy and Precision (CAP). Confidence *Traits* include reliability, manageability, changeability and timeliness; Accuracy *Traits* include data fidelity, error management and procedure robustness; and the Precision *Traits* include resolution, proximity and inclusion zone. Traits are further broken down to atomic *Elements* with properties that can be estimated or measured.

C. Estimating Credibility Intensity and Policy

Observations are drawn from a number of data sources, including the request details (e.g. destination, network access), stored previous records, positioning services, concurrent login activities, and so on. Enterprise sources or hosted services often have suppliers’ specifications for Meantime-Between-Failure (MBF) or 24/7 Availability. Other properties, such as

auditability or retrieval frequency are estimated by IT personnel. This process of establishing credibility rates per source is an off-line procedure, which only needs to be revisited periodically or when sources have changed.

When sources are interrogated, dynamic readings are gathered that identify the active status of an observation. Some observations are binary, i.e. just true or false. Others have an associated degree of strength, i.e. *Level of Intensity (LoI)*. Intensity does not show how credible an assertion is, but measures the strength of the observation, such as the length of ‘Long Duration’ or the proximity to the ‘busy-hour’. Unlike credibility, LoI is obtained dynamically and is specific to the service request. LoI grading is expressed via special scales, where subjective grades (high, medium, low) are transformed by fuzzy indices to unit-less numerals.

While Credibility rates determine the degree of trust in the evidence, *Policy* is used to inject the organization’s own preferences and rules. Policy does not identify *which* attribute prevails, but used to decide *what to do* about it. Policy should not be confused with *impact weighting* that is used to align contributing elements by their significance. Policy prioritization is applied *after* attributes are evaluated, to distinguish between *credible context* and *prioritized context*.

Hence, observations scores are compiled from the inherent value of the aggregated Credibility of all the contributing sources, together with the dynamically observed *Level of Intensity*. When the observations are aggregated into attributes and in-turn into *Key-Factor (KF)* classes, policy-based prioritization that is specific to each profile is applied to the attributes and KFs, producing a score for each profile type. Figure 3 shows context profiles that are compiled from sources (instigating, supporting and qualifying), and rated by their credibility, observed intensity and customized policy.

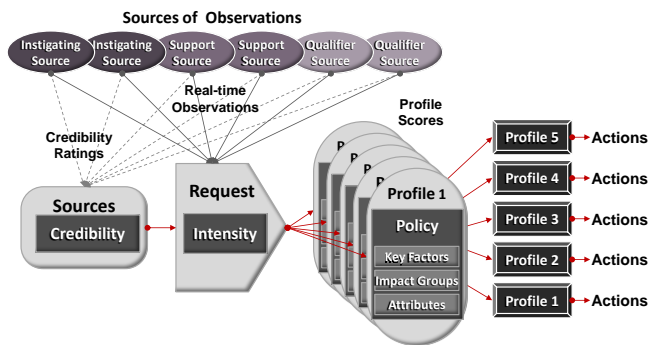


Figure 3. The Credibility Approach to Profiling Requests

D. Data Sources Selection

Observations are based on one or more sources of data. Readings are obtained from sources, such as server logs, historical database, and WLAN logins. These sources are classed as ‘*instigating*’- main sources that identify a triggering fact; ‘*supporting*’- secondary information that confirms or disputes the fact; or ‘*qualifying*’- conditional tables that match the assertions. *Qualifiers* can be special filtering tables, but also records from other systems (e.g. user profile records or work schedules). Raw information becomes *Observation* when it is supported, qualified and the *LoI* is gauged.

The reliability and trustiness properties of all the contributing sources are incorporated in the credibility score of the observation. This combination of sources characterizes the observation and is documented [14]. However, it may be modified by *Probability Logic* [36] conditions, e.g. when sources are temporarily unavailable. While instigating sources may be shared by observations, the particular source combination is uniquely defined for a particular observation, and this contributes to well-distinguished inherent values that are needed for differentiating observations. Figure 4 shows the relationships of sources, observations, attributes and key factors. Observations support assertions that define specific attributes. Attributes are classified by key-factors that describe the perspective of the attributes, e.g. space, time or activity [8]. Prioritized attributes and key-factors are aggregated to produce profile scores, so sources' credibility has considerable influence on the outcome of the model.

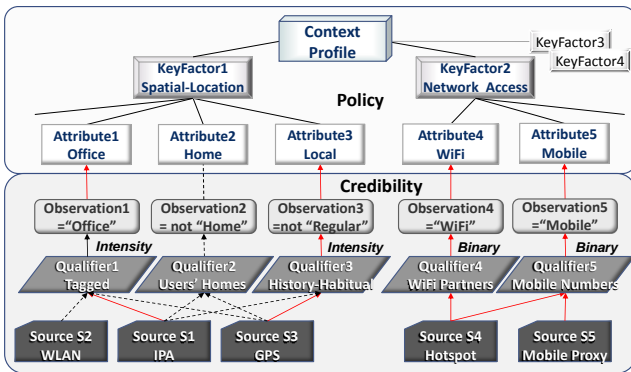


Figure 4. Relationships of Profiles and Credibility & Sources

III. AGGREGATION METHODS FOR A CREDIBILITY MODEL

A. Multi-Criteria Decision Making Methods

Credibility based models rely heavily on aggregation of criteria at several levels: joining estimates of credibility components, combining source credibility into observations, aggregating observations into asserted attributes, corroborating attributes in their key-factors and totaling prioritized key-factors into profiles. Aggregation methods, such as SAW (Simple Additive Weighting) and WPM (Weighted Product Model), which are reviewed in [15], are commonplace but not necessarily suitable for certain context aggregation types. They join the variable members by rank or by proportions, and adjust them with weights. SAW and WPM differ by the type of computing operations: sums (SAW) versus multiplications (WPM). While sums require using the same units, multiplications are unit-less, but need pre-processing to discard zero values. SAW is intuitive but produces linear unscaled results. WPM exaggerates the impact of the weights, which are used as exponents, and produces disproportional scores.

Ranking is used to emphasize higher valued members at the expense of lower ones, and renders the impact of the least significant members (especially in larger sets) almost negligible. OWA (Ordered Weighting Aggregation) is a family of ranking methods [9,10,11]. OWA multiplies the members' values by their transposed ranking order, so that the highest amount is multiplied by the highest ranking number. Induced OWA enables an associated vector to 'induce' ranking, instead

of using the rank numbers as weights. An example of using these methods for combining sources credibility is given in (1). The observation score (OB_b) is compiled from the credibility rates of sources (S_s^{OB}) that are assigned to this observation and their observed level of intensity (LoI_b), weighted by their significance rate (SW_s).

MCDM for Joining source credibility per assertion (1)

$$\forall (Scred|OB_b) > 0 \rightarrow S^{OB} = \{S_{s=1}^{OB}, \dots, S_{s=s'}^{OB}\}, \{SW_{i=1}, \dots, SW_{i=i'}\}$$

$$OB_b = \prod_{j=1}^{j'} (S_s^{OB} \cdot LoI_s)^{SW_i} \quad \text{WPM (a)}$$

$$OB_b = \sum_{s=1}^{s'} (S_s^{OB} \cdot LoI_s \cdot SW_i) \quad \text{SAW (b)}$$

$$S_{s-1}^{OB} > S_s^{OB}, \quad OB_b = \sum_{s=1}^{s'} (S_s^{OB} \cdot SW_{s=i}) \frac{s}{s'} \quad \text{OWA (c)}$$

$$S_{s-1}^{OB} > S_s^{OB}, \quad OB_b = \sum_{s=1}^{s'} (S_s^{OB} \cdot i) \quad \text{I-OWA (d)}$$

B. The Corroboration Algorithm

For proportional corroborative aggregation of members, a new algorithm *CEDAR* (*Corroborative Evidential Diminishing Aggregation Ranking*) is proposed, which is subject to a further study. Cedar uses the ranking order as the processing sequence that progressively diminishes the lesser members' contributions. The diminishing effect is achieved by a coefficient that is calculated from the *residual interval* (1-previous contributions). This coefficient is multiplied by the product of the member's inherent Credibility-based value, and the assigned impact weight. At each step, the residual is reduced in proportion to the contribution, as in (2).

$$\forall f(S_s^{OB}) = SW_{i-1} S_{s-1}^{OB} + SW_i S_s^{OB} (1 - SW_{i-1} | S_{s-1}^{OB}) \quad \text{Cedar (2)}$$

$$OB_{cred_b} = \begin{cases} f(S_s^{OB}) = 0 & s = 0 \\ f((S_{s-1}^{OB}) + SW_i S_s^{OB} (1 - f(S_{s-1}^{OB}))) & s > 1 \end{cases}$$

C. Comparing aggregation Results

Analysis of several aggregation methods shows that there is a great disparity between the produced scores and their distributions of rates, even after normalization into the same scale. In Table I, four different methods aggregate the credibility of the three sources (Instigating, Supporting, Qualifying), which have been already weighted.

TABLE I. COMPARING OBSERVATION RATING AGGREGATION							
Risks	Instigate	Support	Qualify	Ave	SAW	OWA	Cedar
Non Habitual Location	0.365	0.283	0.034	0.227	0.682	0.653	0.560
Undesirable Activity	0.404	0.191	0.053	0.216	0.648	0.516	0.543
Repeated Requests	0.442	0.283	0.032	0.252	0.756	0.781	0.612
Exceeded Authority	0.305	0.053	0.032	0.179	0.358	0.663	0.342
Large data/long duration	0.266	0.032	0.032	0.149	0.298	0.563	0.289
Conflicted Location	0.365	0.311	0.034	0.237	0.711	0.717	0.578
Repeated Failed Authen.	0.305	0.283	0.205	0.264	0.792	0.721	0.603
Repeated DB Hits	0.266	0.032	0.032	0.149	0.298	0.563	0.289
Excessive Net Usage	0.404	0.283	0.032	0.343	0.687	1.091	0.572
Implausible Location	0.365	0.311	0.034	0.237	0.711	0.717	0.578
Unauthorised Updates	0.266	0.205	0.032	0.235	0.470	0.736	0.416
Demanding Media	0.404	0.053	0.032	0.229	0.457	0.861	0.436

Averaging always results in a lower value than the maximum member. This may be acceptable in source credibility aggregation, but in the case of attributes aggregation where credibility should be cumulative (e.g. Urgency KF), or where the largest member (the prime) has a particular role (e.g. Spatial KF), corroborative process should augment the overall

credibility, as in Cedar. OWA increases the score unreasonably when a small contributory member is added to a large one (as highlighted in blue). SAW raises the amounts linearly, and can exceed the scale limits, so further normalization process is required. Cedar performs best, producing consistent, proportional scores within the scale.

IV. COMPUTING CREDIBILITY-BASED SCORES

A. Balancing Credibility Estimates and Consistency

Sources credibility is estimated by the enterprise experts. These subjective estimates may be biased, clustered, or equal-rated, preventing clear differentiation. Hence, it is important to balance estimates using a formal tool, such as AHP (Analytic Hierarchy Process) [19]. AHP applies pairwise comparison to the list of variables and calculates a Consistency Ratio, as in Figure 5. In this example, weights are calculated for the Confidence Measure. Each pair of the (g) elements is assessed pairwise for relative impact. The eigenvector is the g th root of the product of each matrix row, which is then normalized. The resulting eigenvector is the element's weighting rate. To check consistency and randomness, the Consistency Index (CI) is computed from the original estimates and their eigenvectors, and then the Random Index (RI) constant (as observed by Saaty's experiments) is used for the CR (Consistency Ratio), which must remain below one.

AHP calculation of eigenvector for Confidence Weightings												
Avail/Maximu	Avail	High	Auto-Update	Regional	Not Depend.	Fresh Data	Stable Data	Retrieval	Caching	Calc	Log rtm	Expo
Avail 24/7	1.00	2.20	3.00	3.00	1.50	0.90	1.70	1.45	1.10	72.48	0.61	1.84
High MBF	0.45	1.00	0.60	0.50	1.80	0.70	0.60	1.60	0.80	0.13	-0.29	0.75
Auto-Update	0.33	1.67	1.00	2.20	2.75	3.20	1.50	1.85	1.10	32.83	0.50	1.65
Regular Updates	0.67	2.00	0.45	1.00	1.75	0.60	0.40	1.50	0.33	0.13	-0.30	0.74
Not Depend.	1.11	0.56	0.36	0.57	1.00	0.11	0.08	0.50	0.60	0.00	-1.14	0.32
Fresh Data	0.45	1.43	0.31	1.67	9.09	1.00	2.50	3.00	1.80	41.51	0.53	1.70
Stable Data	0.59	1.67	0.67	2.50	12.50	0.40	1.00	1.45	1.80	21.32	0.44	1.55
Retrieval	0.69	0.63	0.54	0.67	2.00	0.33	0.69	1.00	0.90	0.06	-0.39	0.68
Caching	0.91	1.25	0.91	3.03	1.67	0.56	0.56	0.03	1.00	0.04	-0.45	0.64
Calculating Consistency Ratio										9.87	1.00	
Avail 24/7	0.19	0.17	0.50	0.23	0.05	0.16	0.27	0.10	0.07	1.72	9.21	1
High MBF	0.08	0.08	0.10	0.04	0.06	0.12	0.09	0.11	0.05	0.73	9.66	2
Auto-Update	0.06	0.13	0.17	0.17	0.09	0.55	0.24	0.13	0.07	1.60	9.56	3
Regular Updates	0.12	0.15	0.08	0.08	0.06	0.10	0.06	0.10	0.02	0.77	10.27	4
Not Depend.	0.21	0.04	0.06	0.04	0.03	0.02	0.39	0.03	0.04	0.49	15.15	5
Fresh Data	0.08	0.11	0.05	0.13	0.29	0.17	0.39	0.21	0.12	1.55	8.99	6
Stable Data	0.11	0.13	0.11	0.19	0.40	0.07	0.15	0.10	0.12	1.38	8.81	7
Retrieval	0.13	0.05	0.09	0.05	0.06	0.06	0.11	0.07	0.06	0.67	9.84	8
Caching	0.17	0.09	0.15	0.23	0.05	0.10	0.09	0.00	0.06	0.95	14.70	9
CI=Consistency Index										0.2111	0.1099	1.0000
CR=Consistency Ratio										0.150	0.1099	1.0000
Elements: n=9										0.2111	0.1099	1.0000
λmax=n/(n-1) = 0.2111 = CI										0.150	0.1099	1.0000
CR=0.150 = CI/RI for n-1										0.150	0.1099	1.0000
										Observed		

Figure 5. AHP calculation for Confidence

B. Computing Source Credibility

Source credibility is computed from Elements (E_e), Traits (T) and Measures (M_m), which are weighted by E_{w_e} , T_{w_t} , M_{w_m} respectively. Source credibility ($Screds$) is the combined weighted components (using SAW in this case), as in (3).

$$\text{Source } S: 1 \leq s \leq s', \text{ Measure } M: 1 \leq m \leq m' \quad (3)$$

$$\text{Trait } T: 1 \leq t \leq t' \text{ Elements } E: 1 \leq e \leq e',$$

$$\forall (E|S) \neq 0 \quad 1 \leq j \leq j', \{E_1, \dots, E_{j'}\} \text{ Non-zero } (j) \text{ elements}$$

$$\sum_1^{m'} MW_m = \sum_1^{t'} TW = \sum_1^{e'} EW = 1 \quad \text{Weights}$$

$$M_m = TW_{etm} \sum_1^{e'} E_{etm} \cdot (EW_{etm} \cdot \frac{e'_t}{j'}) \quad \text{Normalizing non-zero}$$

$$Scred_s = \sum_{m=1}^{m'} M_{ms} \cdot M_{w_{ms}} \quad \text{Source-dependent Weights}$$

$$\sum_1^{m'} KFMW_{mk} = 1, \quad KF: 1 \leq k \leq k' \quad \text{KF-dependent weight}$$

$$Scred_{sk} = \sum_1^{m'} KFMW_{mk} (TW_{etms} \sum_1^{e'} E_{etms} (EW_{etms} \cdot \frac{e'_t}{j'}))$$

Measures have varying significance in different Key Factors (KF), e.g. Media KF is not affected by Precision, but

Precision is much more important to Spatial KF than Accuracy. Therefore, there is also a key-factor based Measure weighting ($KFMW_{mk}$) that relates to the key-factor identity (k). Figure 6 shows the varying proportions of Measures in sources. For example the dominant component in 'Appointment' is precision, but App&Data, Email and LAN/WLAN login servers are characterized by high accuracy. Confidence is never the dominant component, which may indicate that estimates need to be re-checked.

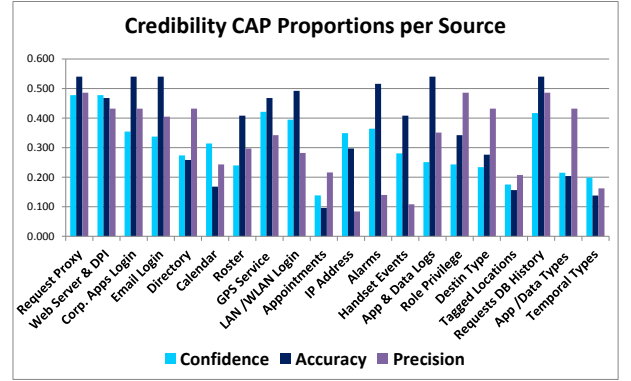


Figure 6. Proportions of CAP Credibility Measures in Sources

C. Transformation of Intensity

Credibility-based models must also consider the variable levels of intensity (LoI) of observed data. Some attributes are always on/off, i.e. binary intensity, but others, e.g. location or integrity, have greater uncertainty that is expressed as intensity grades. Transformation normalizes binary and graded values into unit-less relative scales, ensuring that the score distribution is as wide as possible, for better differentiation. One transformation option is to align LoI_b Min/Max of the Observation (OB_b) with [1,0] as the new limits. This results in *stretching* graded LoI between 0 and 1, producing duplicates of $L=0$ and $H=1$. A better alternative is to determine the high/low limits (excluding 1 and 0) across the whole service request (RQ), so that binary rates are *squeezed* down into the Min/Max scale of all non-binary observations, as in (4). Normalizing across the whole request involves longer real time computation, but maintains better scaling balance.

$$\overline{LoI}_b = \text{Transformed Intensity } LoI_b \quad (4)$$

$$NuRange^{RQ} = [L^{RQ}, H^{RQ}] \rightarrow \text{Per Request } RQ,$$

$$\forall OB_b \in A_a, A_a \in KF_k \rightarrow L^{RQ} = MinLoI_{bak} = \min(OB_{bak}, \dots, OB_{b'a'k'})$$

$$\forall OB_b \in A_a, A_a \in KF_k \rightarrow H^{RQ} = MaxLoI_{bak} = \max(OB_{bak}, \dots, OB_{b'a'k'})$$

$$\overline{LoI}_b = (MaxLoI_{bak} - MinLoI_{bak}) \frac{LoI_b - LoI_{min_{ba}}}{LoI_{max_{ba}} - LoI_{min_{ba}}} + MinLoI_{bak}$$

D. Joining Sources' Credibility Rates per Observation

The observation score is aggregated from the credibility rates of all the contributing sources and qualifiers, weighted according to their perceived significance. The numbers of participating sources vary from one observation to another. For example, the 'Habitual' assertion is instigated by a GPS positioning and is qualified by tagged locations, while supporting information is provided by the historical database that determines the intensity level of the 'habituality'. An *instigating* source is always present, and most observations

require a *qualifying* source, but they don't always have *supporting* information. When combining sources credibility into an observation rate, the sources are weighted according to their types, where instigating sources are given higher impact than supporting or qualifying sources. **Figure 7** shows sources that are: X=instigating, &=supporting or Q=qualifying.

Joint Credibility for Risk Assertions	Risk Element	Request Proxy	Web Server & DPI	Corp. Apps Login	GPS Service	LAN /WLAN Login	App & Data Logs	Role Privilege	Tagged Locations	Requests DB History	App /Data Types
Calculated Credibility:		0.491	0.449	0.435	0.406	0.389	0.379	0.341	0.172	0.471	0.266
Designation Rating:		0.900	0.900	0.700	0.900	0.800	0.700	0.600	0.200	0.600	0.200
Weighted Source Credibility:		0.442	0.404	0.305	0.365	0.311	0.266	0.205	0.034	0.283	0.053
Non Habitual Location	0.201			X					Q	&	
Undesirable Activity	0.216	X									
Repeated Requests	0.252		X								Q
Exceeded Authority	0.255	X								&	
Large data/long duration	0.205			X				Q			
Conflicted Location	0.209		X							&	Q
Repeated Failed Auth	0.264	X		&							
Repeated DB Hits	0.149						X				
Excessive Net Usage	0.343				X	&			Q		
Implausible Location	0.209	X						Q			
Unauthorised Updates	0.174			X				Q		&	
Demanding Media	0.229						X			&	

Figure 7. Joining Credibility per observations for assertion types

E. Attributes Aggregation from Observations

An attributes (A_a) incorporates all observations that support its assertion. The most significant observation (OB_b) is the 'prime', with the highest value of combined sources' *Scred* and the observed *Lol*. Cedar gives precedence to the most significant observation and minimizes the impact of lower rated ones. It ensures that scores remain within the scale without having to normalize them, as in (5).

$$\begin{aligned}
 &LoI_b \ \& OB_b: 1 \leq b \leq b', A: 1 \leq s \leq s' \text{ Aggregate Attributes (5)} \\
 &OB_b = \sum_{s=1}^{s'} (Scred_{sb} \cdot LoI_b \cdot Sw_s) \text{ Joint Source Cred \& LoI} \\
 &A_a = \sum_{b=1}^{b'} OB_b \text{ SAW} \\
 &f(OB_b) = OB_{b-1} + OB_b(1 - |OB_{b-1}|), \quad OB_{b-1} > OB_b \\
 &A_a = \begin{cases} OB_0 = 0 & b = 0 \\ ((f(OB_{b-1}) + OB_b(1 - f(OB_{b-1}))) & b > 1 \end{cases} \text{ Cedar}
 \end{aligned}$$

F. Key-Factors Aggregation from Attributes

Key Factor (KF_k) aggregation depends on their type. Some factors require a single attribute selection, while others are cumulative, with a selected 'prime' attribute (A_p). The prime determines the prioritization that dictates the outcome, e.g. prime=Home means that home-working policies are applied, but prime=Office means that different rules prevail. In cases of cumulative key-factors with contributive evidence, a corroborative aggregation method is required. The attributes and key-factors are also weighted by the policy-based prioritization rates, for attributes (AW_a) and factors (KFW_k) before adding up the profiles scores, as in (6).

$$\begin{aligned}
 &RCP C: 1 \leq c \leq c', A: 1 \leq a \leq a' \text{ Aggregating Factors (6)} \\
 &KeyFactor \ KF_k := \{A_{1k}, \dots, A_{a'k}\}, \quad 1 \leq k \leq k', \quad A_p = \max(A_a) \\
 &f(A_a W_a) = A_{a-1} W_{a-1} + A_a W_a (1 - A_{a-1} W_{a-1}) \text{ Cedar} \\
 &KF_k = \begin{cases} f(A_a W_a) = 0 & a = 0 \\ ((A_{a-1} W_{a-1}) + A_a W_a (1 - f(A_{a-1} W_{a-1}))) & a > 1 \end{cases} \\
 &RCP_c = \sum_1^{k'} KF_k \cdot KFW_k \text{ Context Profile}
 \end{aligned}$$

V. ALTERNATIVE APPROACHES

A. Entropy for Component Performance

Information Theory and *Entropy* offer insights for parameters inter-dependence and the level of doubt associated with the estimated values. Uncertainty is intrinsic in behavioral models, so quantifying tools are particularly useful. Concepts of Mutual Information and Conditional Mutual Information can be adapted to assess the model's components, with extended applicability to arrays and random distributions, as proven in [2] and applied in [3]. In (7), Context Profiles Efficiency (*CPE*) is quantified, based on attribute credibility rate ($Acred$) and its probable occurrence rate per profile type ($P(A_i)$) in the *RQS* database of proven requests. Context Profile (CP_c) is treated as a vector of all attributes (A_i) with n members. The attributes entropy rates are summed up and normalized, to provide the profile's entropy.

$$\begin{aligned}
 &H(CP_c) = exp[-\log_2(P(CP_c))] \text{ Entropy Information (7)} \\
 &H(CP_c) = -\sum_1^n P(A_i) \log_2(P(A_i)) \text{ CP is attribute vector} \\
 &A_{1 \leq i \leq n} = Acred_i \cdot P(A_i | RQS) \text{ Combined Cred/Occurrence} \\
 &Entropy for an array of attributes: (A_1, A_2, A_3) = \\
 &-\sum_{i=1} \sum_{i=2} \sum_{i=3} P(A_1, A_2, A_3) \log_2(P(A_1, A_2, A_3)) \\
 &CPE(CP_c) = -\sum_1^n P(A_i) \cdot \log_2(P(A_i)) \frac{1}{\log_2(n)} \text{ Normalized average}
 \end{aligned}$$

Entropy can be used to indicate the profiles' performance and assess attribute impact on the profile scores. It can indicate where additional assurance is required, perhaps by adding supporting sources. The relative Entropy scores can determine which profile type prevails in cases of a tie (equal profile scores), so that the profile with the lowest uncertainty is selected. Entropy was computed for five profiles in the Risk Model, with 24 risks. The Entropy procedure was adapted for using combined attribute occurrence rate with credibility rate. Entropy was calculated for each attribute and the entropy for each profile type was aggregated, as shown in **Figure 8**.

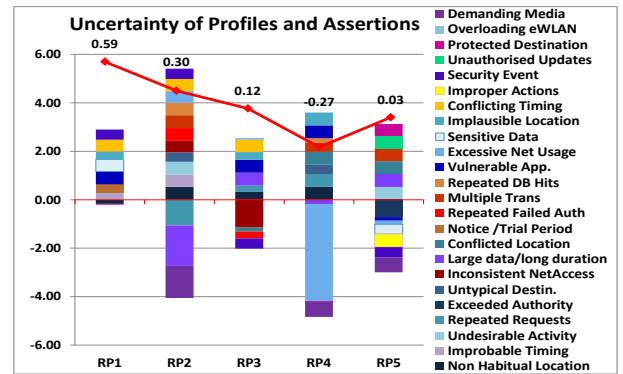


Figure 8. Entropy of risk profiles

This example shows that RP4 (Risk Profile 4) has the lowest Entropy by a good margin. The 'Excessive Net Usage' risk (light blue) dominates RP4, but its Entropy is lower than most other risks, so trustiness is high. However, such high dependence on a single risk is inadvisable, calling for further examination of the key-factor structure. The span of Entropy in RP1 is significantly smaller than RP2 or RP4, and many risks

are above zero, i.e. high uncertainty, so RP1 is not as safe a verdict as it should be.

B. Belief Function for Joint Credibility

The DST Belief function is considered as an alternative to probability models as well as to the Credibility model. DST assigns Belief ‘masses’ to the set of options, where Plausibility is whatever has not been assigned as a contrary belief. This allows for some unknown mass to be included in the plausibility. DST was widely criticized as inconsistent and paradoxical, and was proven unable to deal with large conflict. Despite the criticism, the DST debate has highlighted some important principles that are glossed over by the probability approach. DST can support equal-probability and partial knowledge. It separates the *degree of support* (=Level of Intensity) from the *source reliability* (=Credibility), as in this paper. The DST combination rule, as in (8), focuses on the intersection of sets, using Exclusive-OR to combine ‘masses’. This segregates conflicting evidence as contrasting data sets, but they are used merely as normalization [7]. Thus, conflict is discarded instead of taken it into account, leading to counter-intuitive results. A significant drawback of the rule is the complexity when joining more than two sources, because DST defines ‘mass’ as a *power set* function, that quickly ramps up the number of combinations that need to be computed.

$$m_{1,2}(N) = m_{1,2}(N) = (m_1 \oplus m_2)(N) \quad \text{The Dempster Rule (8)}$$

$$m_{1,2}(A) = \frac{\sum_{B \cap C = A, A \neq \emptyset} m_1(B)m_2(C)}{1 - \sum_{B \cap C = \emptyset} m_1(B)m_2(C)}, \quad \text{where } S = B \cup C \rightarrow$$

$$m_{1,2}(B) = \frac{m_1(B)m_2(B) + m_1(B)m_2(S) + m_1(S)m_2(B)}{1 - (m_1(B)m_2(C) + m_1(C)m_2(B))} \quad \text{Expanded}$$

Table II compares handling of concordant and discordant attributes using the Dempster rule, Cedar, SAW and OWA. It is assumed that credibility/intensity values represent the masses for DST and weighted values for SAW. The scenario data includes two pairs, with discordant (in red), and concordant (in blue) contributing attributes. Mass1 and Mass2 are computed from their respective credibility and LoI values, taken from the model scenarios and source credibility calculation.

		LoI1	Cred	Mass1	LoI2	Cred2	Mass2	DST	Cedar	SAW	OWA
1	Office	0.800	0.776	0.621	0.300	0.617	0.185	0.317	0.192	-0.004	0.147
	-Home Set-Cred	0.500	0.776	0.388	0.700	0.617	0.432				
2	Office	0.800	0.776	0.621	0.300	0.617	0.185	0.209	0.878	0.406	0.380
	Office Set-Cred	0.500	0.776	0.388	0.700	0.617	0.432				
3	Branch	0.800	0.617	0.494	0.300	0.582	0.175	0.204	0.193	-0.028	0.110
	-Regular Set-Cred	0.700	0.617	0.432	0.600	0.582	0.349				
4	Branch	0.800	0.617	0.621	0.300	0.582	0.175	0.188	0.875	0.383	0.363
	Branch Set-Cred	0.500	0.617	0.388	0.600	0.582	0.349				
	Span							0.129	0.686	0.435	0.271

The results show that DST produces higher rates when conflict is present (example 1) than without conflict (example 2), which is counter-intuitive, and has the lowest span of scores, i.e. low differentiation. SAW produces negative results, which is a disadvantage. OWA shows disproportionately lower or higher values compared with Mass1 and Mass2 (example

3,4). By contrast, Cedar provides consistently proportional, positive scores, in the widest spread.

C. The Bayes Classifier

The most popular alternative to Credibility-based modelling is probability-based classification. The Bayes classifier is used to determine the Context Profile type (CP_c) per new request with a set of attributes $\{A_1, \dots, A_a\}$. A request is treated as a vector of all its attributes, irrespectively of any key-factors classification, since the probabilities of these attributes are not dependent on KF prioritization. The probability of an attribute is the occurrence rate within each profile type in the RQS dataset of requests that have been computed using the credibility-based model. A profile PC_c is classified by Bayes classifier in (9).

$$CP: 0 \leq c \leq c', \quad A: 0 \leq a \leq a' \quad \text{Bayes theorem (9)}$$

$$ReqA = \{A_{i=1}^{n=k'a'}\} = A_{11}, \dots, A_{1a'}, \dots, A_{k1}, \dots, A_{n=k'a'} \quad \text{No KF}$$

$$p(CP_c|A_i^n) = \frac{p(CP_c|RQS) p(A_i^n|CP_c)}{(A_i^n|RQS)}$$

$$p(CP_c|A_i^n) \propto p(CP_c|RQS) \cdot p(A_1, \dots, A_n|CP_c)$$

$$p(A_1 \dots A_n|CP_c) = p(A_1|CP_c)p(A_2|CP_c) \dots p(A_n|CP_c) \quad \text{Chain rule}$$

$$c = \underset{x}{\text{argmax}} p(CP_c|A_i^n) \cdot \prod_i p(A_i^n|CP_c) \quad \text{Classify Profile}$$

The Bayes results for the Risk Model, with 24 attributes and 60 scenarios were respectable, with overall 75% correct profile predictions, but the Credibility model outperforms Bayes with 95%, as shown in Table III. When classifying Business Context profiles with 56 attributes and 50 scenarios, only 50% of Bayes predictions were correct - no better than pure chance. This confirms that Bayes requires an unrealistically large training data to provide sufficiently high ratio of data-points to variables. To quote MYCIN developers [1], they rejected Bayes because it would require “unfeasibly large numbers of ‘proven’ cases” while “the assumption of independence is unrealistic”.

	Model	Bayes
Non-marginal	100.00%	72.73%
Marginal	88.89%	77.78%
Overall	95.00%	75.00%

VI. MODEL COMPONENTS ANALYSIS

Different compositions of model components were tested, to gauge their effectiveness. The ‘Full Model’, containing Credibility, LoI, and Policy (by KF, Attributes and Groups), represents the expected results, against which deviations (wrong profile type, increased ambiguity, equal-ranking and blurring) are noted. The total deviation in Figure 9a rises sharply when Policy components are omitted. The spread of scores is most affected by the lack of Attribute Policy, though the standard deviation and the mean remain generally steady. Compositions without Attribute Policy are particularly prone to equal-ranking, as the trend line shows.

In Figure 9b, analysis per scenario is shown, with score distortions (e.g. wrong profile types) noted where peaks are not synchronized with the Full Model (in red). It is evident that No-Credibility and No-Intensity graphs (blue and purple) track

the model graph, while No-Policy (light green) and LoI-only (black) are much flatter, with inferior score differentiation. In Figure 9c, ambiguity per scenario is shown, measured by the 1st margins (between largest and second-largest scores). The *Max1stMargin* trend line (black) tracks the Full Model (red) very well, while other compositions are erratic.

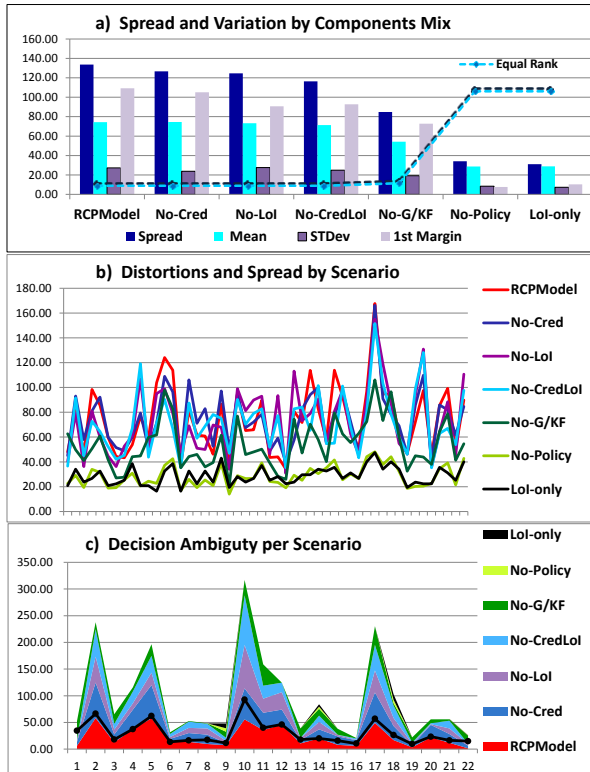


Figure 9. a) Spread & Variation b) Score Deviation c) Score Ambiguity

This analysis reveals that Credibility and LoI support consistent choice of profiles and LoI increases the distribution span. Policy by Groups/Key-Factors increases differentiation and reduces ambiguity, while Attributes' Policy avoids equal-ranking and inconclusive verdicts. None of the components is superfluous, as they all have significant impact.

VII. RELATED RESEARCH

Although the need to account for accuracy of context data has been recognized in many papers, it is rarely integrated into decision making algorithms. In [24], lack of data quality is proven detrimental to decision making, due to increased task complexity. The Credibility concepts appear in various disciplines: metrology [4], actuarial risk [5] and medicine [1]. Data inconsistency issues are raised in [12], noting that OWL (Web Ontology Language) lacks such consideration, but *Accuracy* is now included in enhanced XCML (eXtended Context Modeling Language) [14]. In [31], quality of context in terms of privacy, security, resolution, completeness, and precision, are noted using OWL-DL semantics, still not covering the whole range of credibility properties. In [18], 'credibility' in MCDM (Multi-Criteria Decision Making) has been linked with approximate (fuzzy) reasoning, but no indication of how it is ascertained.

On Credibility aggregation mechanisms, MCDM methods such as SAW (Simple Additive Weighting) and WPM (Weighted Product Model), are reviewed in [15]. They are lightweight and fast, but ignore ranking and primacy. Where aggregation by ranking is called for, OWA (Ordered Weighting Aggregation) is popular, as in [9, 10], and expanded in [11]. However, OWA produces disproportional scores that do not corroborate the prime member in proportion to each contribution and do not remain within a given scale.

The most common alternative to the Credibility-based model is using probability-based models, especially Bayesian classifiers, as described in [29]. Extended Bayesian algorithm is used for context measured by environmental sensors in [16], which groups together dependencies, to assess 'correctness'. In [27], Naïve Bayes algorithm outperforms the proposed extensions, but [28] claims success with an augmented classifier that copes with dependencies, using Galois lattice. Independence of context attributes is not achievable, and methods that rely on such independence will distort the outcome. For a substantial number of variables (above ten), Bayes requires a large proven data set that represents all the combinations that may occur, which is unfeasible for behavioral context.

The method of DST '*Belief Function*' tackles uncertainty of evidence. In [23], the Dempster-Shafer Theory (DST) is used to assess sensors' *dissimilarity*. In [17], DST is proven for intervals of probabilities, but paradoxes are still ignored. In [32], Shafer admits that Belief Function is best used for an interactive subjective process. However, DST offers some advantages: belief functions do not require prior probabilities and missing data can be estimated under 'Plausibility'. As per [33], belief function models do not deteriorate as fast as probability models, when the gap between estimations and reality grows. However, the Dempster combination rule as detailed in [7], cannot cope with conflict. The rule is extended in [26], by assigning separate masses to conflict, but this eliminates discord, not integrate it.

Quality-of-Information is measured in [6,22] for the reliability of *unknown* sources. For the Credibility approach, knowledge of the sources is essential. It requires structuring attributes from their sources with a bottom-up approach, as in [25], where attribute trees are built up from observations that link to sources properties. Observations must allow for both concordant and discordant observations, as described in [13], to account for discord and reduce the score accordingly.

Several tools are useful help in different parts of the model. The credibility components estimates and the assigned weights must be non-ransom, balanced and well spread. Saaty's AHP (Analytic Hierarchy Process), as explained in [19] provides a tool to do that. Information Theory and Entropy help to assess context profiles uncertainty. In [3] Entropy and Conditional Mutual Information are utilized to correlate contributions to service composition quality. The Shannon rule is augmented by [2] to any distribution of random intervals. To manage dynamically context uncertainty, Goertzel's *Probability Logic* [36] and the OpenCog project provide some inspiration, with advanced, but complex, logic functions such as induction, abduction, analogy, speculation, and causality.

VIII. CONCLUSIONS

In this paper, a pragmatic approach is proposed which derives context credibility from the properties of sources and incorporates it in the evaluation algorithms. The credibility based model computes context afresh for every instance, with no prior probabilities or training data, and is not vulnerable to contamination by previous faulty classification. This approach can cope with more complex modelling, higher levels of uncertainty and lack of previous history. Since credibility rates are computed off-line, this method reduces the load on real-time processing. However, such a credibility model is useful only where information is drawn from a diverse range of known sources that have measurable reliability and trustiness, so that credibility rates can be differentiated and ultimately produce good distribution of attribute scores. The Credibility approach incorporates source credibility, intensity and policy in the scoring of criteria, so differentiation is also achieved due to the dynamically gauged Intensity and the variable mix of sources per observation.

The accuracy of the Credibility-based model relies heavily on an appropriate aggregation method that is corroborative and proportional to the contributions. It must aggregate two members as well as numerous contributing members in an equitable manner and produce reliable and conclusive results. The new Cedar algorithm aggregates contributory members in a recursive process. It accrues contributions proportionally to the inherent credibility-based values, in a diminishing ratio, so that less credible members have less impact. This is achieved by using a coefficient that is calculated at each iterative step as the absolute value of the residual interval, after previous contributions have been subtracted. This algorithm, which operates on ordered set of members, always augments the 'prime' (the largest member), with supportive contributions. Discord is accounted for by aggregating conflicting members as negative amounts in the same way, decrementing the score in proportion to the conflicting evidence.

The Credibility approach is proven to be robust and can cope with more elaborate context structures, attributes discord and dependency. Furthermore, incorporating Credibility in the evaluation integrates further knowledge that goes beyond occurrence likelihood and provides a better way for informed decisions to be made.

REFERENCES

- [1] D Heckerman, E Shortliffe "From Certainty Factors to Belief Networks" Artificial Intelligence in Medicine 1992
- [2] R Renner, U Maurer "About the Mutual (Conditional) Information" IEEE 2002
- [3] J Jiang, Q Bai "Correlated Contribution Analysis for Service Composition in Dynamic Environments" IEEE2013
- [4] JCGM 200 "International vocabulary of metrology" 2008
- [5] Actuarial Standards Board ASOP 25 "Credibility Procedures" 2012
- [6] D Wang, L Kaplan, T Abdelzaher, C Aggarwal "On Credibility Estimation Tradeoffs in Assured Social Sensing" IEEE 2013
- [7] K Sentz "Combination of Evidence in Dempster-Shafer Theory"2002
- [8] R Copeland, N Crespi "Classifying and Aggregating Context Attributes for Business Service Requests - No 'One-Size-Fits-All'" in-press SSC IEEE 2014.
- [9] R Yager, D Filev "Induced Ordered Weighted Averaging Operators" IEEE 1999
- [10] R Senge and E Hullermeier "Pattern Trees for Regression and Fuzzy Systems Modeling" IEEE 2011
- [11] Yu Yi, T Fober, E Hullermeier "Fuzzy Operator Trees for Modeling Rating Functions" IJCAI 2009
- [12] C Bettini et al "A Survey of Context Modelling and Reasoning Techniques" Elsevier 2008
- [13] P. Meyer, M. Roubens, "Choice, ranking and sorting in fuzzy multiple criteria decision aid, in Multiple Criteria Decision Analysis: State of the Art Surveys" Springer-Verlag 2005
- [14] R Robinson, K Henriksen "XCMML: A runtime representation for the Context Modelling Language" IEEE 2007
- [15] K.Savitha, Dr.C.Chandrasekar "Vertical Handover decision schemes using SAW and WPM for Network selection in Heterogeneous Wireless Networks" GJCST (2011)
- [16] M Huebscher, J McCann, N Dulay "Fusing multiple sources of context data of the same context type" Xplore 2006
- [17] J Pearl "On Probability Intervals" California University 1988
- [18] R Yager "On Considerations of Credibility of Evidence," International Journal of Approximate Reasoning, 1992
- [19] G Coyle "The analytic hierarchy process (AHP)" Pearson Education 2004
- [20] R Copeland, N Crespi "Controlling enterprise context-based session policy and mapping it to mobile broadband policy rules" in press, IEEE ICIN 2012
- [21] R Copeland, N Crespi "Selecting Access Network for BYOD enterprises with Business Context (eBC) and enterprise-centric ANDSF" Springer, Mobilware 2012
- [22] D Wang, T Abdelzaher, H Ahmadi, J Pasternack, C Aggarwal "On Bayesian Interpretation of Fact-finding in Information Networks" IEEE 2011
- [23] M Florea, E Bosse "Dempster-Shafer Theory: combination of information using contextual knowledge" ISIF 2009
- [24] W Jung et al "An Experimental Study of the Effects of Contextual Data Quality and Task Complexity on Decision Performance" IEEE2005
- [25] R Barros, R Cerri, P Jaskowiak, A de Carvalho "A Bottom-Up Oblique Decision Tree Induction Algorithm" IEEE2011
- [26] F Campos, A Neves, F. Campello de Souza "Decision Making under Subjective Uncertainty", IEEE 2007
- [27] J Sheppard, S Butcher, M. Kaufman, C MacDougall "Not-So-Naïve Bayesian Networks and Unique Identification in Developing Advanced Diagnostics" IEEE 2006
- [28] L Frey, D Fisher "Augmented Naïve Bayesian Model of Classification Learning" 2003
- [29] H Zhang "Augmenting Naive Bayes for Ranking" International Conference on Machine Learning 2005
- [30] Harry Zhang "The Optimality of Naive Bayes" FLAIRS2004 conference 2001
- [31] J Filho, A Miron, I Satoh, J Gensel , H Martin "Modeling and Measuring Quality of Context Information in Pervasive Environments" ICAINA 2010
- [32] G Shafer, J Pearl "Readings in Uncertain Reasoning" Chapter 7 "Belief Functions", Morgan Kaufman Publishers 1990
- [33] P Smets, "Practical Uses of Belief Functions" Universite Libre de Bruxelles, Brussels-Belgium
- [34] Zou Yi, Ho Yeong Khing, Chua Chin Seng, Zhou Xiao Wei "Multi-ultrasonic sensor fusion for autonomous mobile robots"
- [35] G Shafer "Perspectives on the Theory and Practice of Belief Functions" International Journal of Approximate Reasoning 1992
- M Iklé, B Goertzel "Probabilistic Logic for General Intelligence" 2010 wiki.opencog.org/wiki/home/images/3/3c/PLN_Agents_matt_v5.pdf